

## ORIGINAL ARTICLE

# Diffuse reflectance spectroscopy characterises the functional chemistry of soil organic carbon in agricultural soils

Johanna Wetterlind<sup>1</sup>  | Raphael A. Viscarra Rossel<sup>2</sup>  | Markus Steffens<sup>3,4</sup>

<sup>1</sup>Department of Soil and Environment, Swedish University of Agricultural Science, Skara, Sweden

<sup>2</sup>Soil and Landscape Science, School of Molecular and Life Sciences, Curtin University, Perth, Western Australia, Australia

<sup>3</sup>Department of Soil Sciences, Research Institute of Organic Agriculture FiBL, Frick, Switzerland

<sup>4</sup>Lehrstuhl für Bodenkunde, Department für Ökologie und Ökosystemmanagement, Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Technische Universität München, Freising-Weihenstephan, 85350, Germany

## Correspondence

Johanna Wetterlind, Department of Soil and Environment, Swedish University of Agricultural Science, PO Box 234, Skara SE-532 23, Sweden.  
Email: [johanna.wetterlind@slu.se](mailto:johanna.wetterlind@slu.se)

## Funding information

Australian Research Council, Grant/Award Number: DP210100420; Svenska Forskningsrådet Formas, Grant/Award Numbers: 2017-00887, 229-2010-951

## Abstract

Soil organic carbon (SOC) originates from a complex mixture of organic materials, and to better understand its role in soil functions, one must characterise its chemical composition. However, current methods, such as solid-state <sup>13</sup>C nuclear magnetic resonance (NMR) spectroscopy, are time-consuming and expensive. Diffuse reflectance spectroscopy in the visible, near infrared and mid-infrared regions (vis-NIR: 350–2500 nm; mid-IR: 4000–400 cm<sup>-1</sup>) can also be used to characterise SOC chemistry; however, it is difficult to know the frequencies where the information occurs. Thus, we correlated the C functional groups from the <sup>13</sup>C NMR to the frequencies in the vis-NIR and mid-IR spectra using two methods: (1) 2-dimensional correlations of <sup>13</sup>C NMR spectra and the diffuse reflectance spectra, and (2) modelling the NMR functional C groups with the reflectance spectra using support vector machines (SVM) (validated using 5 times repeated 10-fold cross-validation). For the study, we used 99 mineral soils from the agricultural regions of Sweden. The results show clear correlations between organic functional C groups measured with NMR and specific frequencies in the vis-NIR and mid-IR spectra. While the 2D correlations showed general relationships (mainly related to the total SOC content), analysing the importance of the wavelengths in the SVM models revealed more detail. Generally, models using mid-IR spectra produced slightly better estimates than the vis-NIR. The best estimates were for the alkyl C group ( $R^2 = 0.83$  and  $0.85$ , vis-NIR and mid-IR, respectively), and the O/N-alkyl C group was the most difficult to estimate ( $R^2 = 0.34$  and  $0.38$ , vis-NIR and mid-IR, respectively). Combining <sup>13</sup>C NMR with the cost-effective diffuse reflectance methods could potentially increase the number of measured samples and improve the spatial and temporal characterisation of SOC. However, more studies with a wider range of soil types and land management systems are needed to further evaluate the conditions under which these methods could be used.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *European Journal of Soil Science* published by John Wiley & Sons Ltd on behalf of British Society of Soil Science.

### Highlights

- Diffuse reflectance spectroscopy was used to characterise and model SOC functional chemistry.
- NMR derived C functional groups could be modelled with vis-NIR and mid-IR diffuse reflectance spectra.
- The methods allow for characterisation of SOC chemical composition on whole mineral soil samples.
- The approach can improve the spatial and temporal characterisation of SOC composition.

### KEYWORDS

$^{13}\text{C}$  NMR, C functional groups, C turnover, mid-IR spectroscopy, NIR spectroscopy, soil organic matter composition, soil organic matter quality

## 1 | INTRODUCTION

Soil organic matter consists of a wide range of heterogeneous materials in all stages of decomposition, closely interacting with the soil mineral matrix (Lehmann & Kleber, 2015). To better understand the mechanisms controlling carbon (C) dynamics, we need information on the chemical composition of the soil organic carbon (SOC), the soil physicochemical properties and environmental factors (Kögel-Knabner & Rumpel, 2018; Paré & Bedard-Haughn, 2013; Schmidt et al., 2011; Viscarra Rossel et al., 2019). This study pertains to the characterisation of organic functional groups in SOC.

Solid-state  $^{13}\text{C}$  Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful experimental technique used in different disciplines to elucidate the atomic and molecular structure of a wide range of substances. The main advantages of NMR spectroscopy are that it is non-destructive and the sample can be used for other experiments, solid and liquid samples can be analysed, no extractions are needed and it gives comprehensive and semi-quantitative information on the chemical composition of a sample for one or several selected elements. It is commonly used to quantitatively determine the chemical composition of SOC (Bonanomi et al., 2013), to deduce SOC's degree of decomposition and to allow an estimation of more resistant fractions. Weng et al. (2021) stressed the point that NMR was used to prove and disapprove various theories and hypotheses on SOC dynamics and stabilisation. The technique does not provide the structural organisation of SOC on a molecular level but can broadly differentiate C functional groups (Audette et al., 2021). Most studies identify alkyl C, O/N-alkyl C, aromatic C and carbonyl C groups (Audette et al., 2021; Baldock et al., 1992; Kögel-Knabner, 1997, 2000). Chemical shift ranges can be fitted to four spectral

regions, labelled as (1) alkyl C (10–45 ppm; long-chain polymethylene type structures, for example, fatty acids, waxes and resins); (2) O-alkyl C (45–110 ppm; mostly carbohydrates); (3) aromatic C (110–160 ppm; protonated and C substituted aromatics and unsaturated C and oxygenated aromatics); and (4) carboxyl C (160–200 ppm; carboxylic C, esters and amides) (Baldock et al., 1992; Oades et al., 1987).

Baldock et al. (1997) evaluated the potential of solid-state  $^{13}\text{C}$  NMR spectroscopy to assess the extent of decomposition of natural organic matter. They described a strong link between the progressing decomposition of natural organic matter, a relative increase in alkyl C and relative decrease in O-alkyl C. This can be explained by the characteristic hydrophobicity and more resistant alkyl materials on the one side and the easily decomposable nature of polysaccharides and proteins on the other side. In a recent meta-analysis, Audette et al. (2021) drew a comprehensive summary of the origin and lability of the NMR-derived functional C groups and showed a clear influence on the proportions of the different  $^{13}\text{C}$  NMR-derived C groups of changes in agricultural management practises (i.e., fertilisation, tillage, crop rotation and liming), demonstrating the usefulness of this type of information for guiding agricultural practises and improving soil health.

Due to the small content of total SOC in soil samples and the low natural abundance of the  $^{13}\text{C}$  isotopes, the measurement with NMR can be slow, which limits the number of samples that can be analysed, restricting the use of the method to smaller, dedicated studies (Baldock et al., 1989; Kinchesh et al., 1995). In addition, paramagnetic materials such as iron may interfere with the measurements further reducing the signal-to-noise ratio, so that they have to be removed by treating the samples with

hydrofluoric acid (HF) (Mathes et al., 2002). This treatment results in an increase in SOC content; however, it is also associated with varying degrees of SOC loss, both in terms of total loss and selective loss of organic compounds. This may result in biased interpretation of the SOC chemical composition (Sanderman et al., 2017).

Diffuse reflectance spectroscopy in the visible, near and mid-infrared (vis-NIR: 350–2500 nm or 28,571–4000  $\text{cm}^{-1}$ ; mid-IR: 2500–25,000 nm or 4000–400  $\text{cm}^{-1}$ ) are rapid, non-destructive, methods commonly used in soil science (Soriano-Disla et al., 2014). Reflectance spectra in the mid-IR region are the result of interactions between the radiating energy and the bonds in molecules of soil constituents. In the NIR, the spectra result from overtones and combinations of the fundamental vibrations in the mid-IR region, while in the visible range the primary processes are electronic excitations (Stenberg et al., 2010). The methods provide qualitative information on the fundamental composition of the soil, including clay and iron oxide minerals, organic matter, water and particle size. Hence, that information is the basis for the creation of models from the spectra used to estimate several properties. However, the information in the diffuse reflectance spectra is overlapping and complex, and calibration models are needed for quantitative analysis. Absorbance is bond specific but is also affected by the type of functional group, its neighbouring molecules and hydrogen bonds (Miller, 2001). Information on SOC can be found in several regions of the mid-IR and vis-NIR spectra, and corresponds to, for example,  $-\text{CH}$  and  $-\text{CO}$  groups. SOC content is one of the most commonly modelled and best-predicted soil properties using these techniques (Stenberg et al., 2010). Because the information is related to specific molecular bonds and their surrounding chemistry, SOC content is in fact predicted by its chemical composition, and a number of studies have shown the potential for vis-IR spectroscopy to predict different aspects of the organic matter quality (Knox et al., 2015; Viscarra Rossel & Hicks, 2015). The possible advantage of reflectance spectroscopy for characterising the composition of soil organic C is that the chemical information in a soil sample might be gained from the analysis of the whole soil without C fractionation (e.g., into particulate, mineral associated or pyrogenic organic C) or HF-treatment.

We found a number of studies that explored the relationship between vis-NIR or mid-IR spectra and solid-state  $^{13}\text{C}$  NMR spectra (Forouzangohar et al., 2013, 2015; Kang et al., 2017; Leifeld, 2006; Ludwig et al., 2008; Terhoeven-Urselmans et al., 2006). These studies used HF-treated mineral soils (for both solid-state  $^{13}\text{C}$  NMR and the vis-NIR and mid-IR spectroscopy) or specific soil C fractions, for example, litter, particulate and mineral-associated organic carbon. To our knowledge, there are

no other published studies on the characterisation of soil organic C chemistry with spectroscopy focusing on whole agricultural soils.

Given this research gap, our aim here was to test if vis-NIR and mid-IR diffuse reflectance spectroscopy could characterise the functional chemistry of SOC in whole arable soils with as little pre-treatment as possible, that is, without HF-treatments or soil C fractionation. To do so, we derived 2-dimensional correlations of solid-state  $^{13}\text{C}$  NMR spectra and vis-NIR and mid-IR spectra, compared assignments of the NMR functional organic C groups and the corresponding frequencies in the vis-NIR and mid-IR, and modelled the NMR functional groups with the reflectance spectra using support vector machine regression. For our experiments, we used 99 Swedish soil samples with a wide range in SOC content.

## 2 | MATERIALS AND METHODS

### 2.1 | Soil samples and analyses

We used 99 mineral soil samples from the 0–20 cm layer of agricultural fields in Sweden (Figure 1). The soils were selected from 12,500 soil samples collected in a national campaign run by the Swedish Board of Agriculture in 2010 and 2011 and archived at the Swedish University of Agricultural Sciences. The 12,500 samples were collected in a regular grid of one soil sample per  $\text{km}^2$ , randomly moving the sampling site 1–150 m around the grid node, across about 90% of Swedish agricultural land. The 12,500 samples were air dried, sieved to 2 mm and analysed for soil texture, soil organic matter content (measured as loss on ignition and corrected for structural water in clay [Ekström, 1927]) prior to archiving. Soil texture was divided into clay ( $<0.002$  mm), silt (0.002–0.06 mm) and sand (0.06–2 mm). Clay content was analysed using a sedimentation method modified from Gee and Bauder (1986); the sand fraction was determined by sieving and the silt fraction was determined by difference. The 99 samples used in this study were selected using stratified random sampling to cover a wide range of soil texture and soil organic matter content by dividing the 12,500 soil samples into classes based on soil texture and organic matter content and randomly select samples within those classes (Figures 1, S1 and Table S1). A maximum amount of organic matter was set to 16% to focus on mineral soils and the clay content was limited to 40% because of a focus on more sandy soils in a joint study. The 99 samples selected for this study represent a cross section of the possible variations in soil texture and organic matter content of the 12,500 Swedish

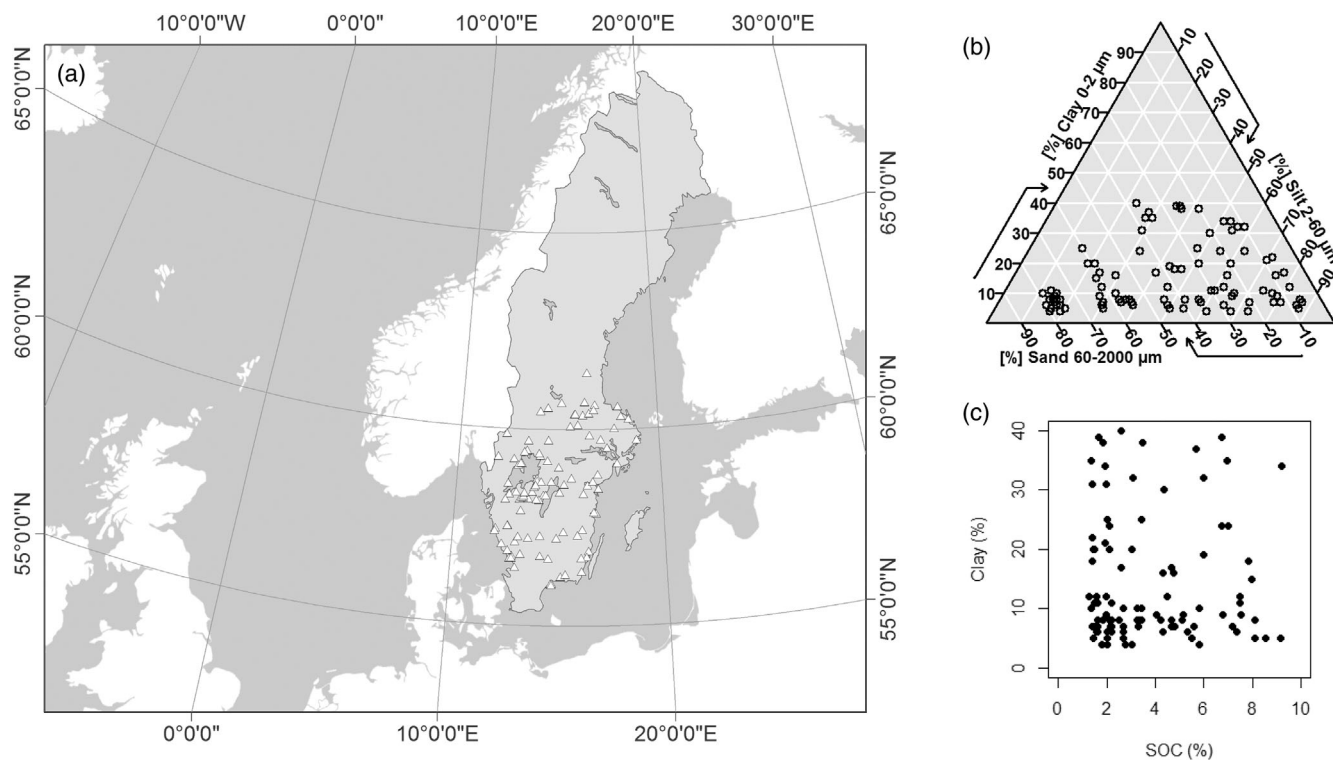


FIGURE 1 (a) Location of the 99 soil samples and (b) soil texture and (c) soil organic carbon content (SOC) in the samples

arable soils (within the set organic matter content and texture boundaries). For comparison with the analyses of the functional C groups, the 99 samples were also analysed for SOC. There were no carbonates present in the soil samples and the 99 samples were analysed for SOC through dry combustion on an EuroEA elemental analyser (Hekatech GmbH, Wegberg, Germany). Swedish soil is young (i.e., mainly formed during the quarternary period) and strongly affected by processes during and after the last glacial period (Karlsson et al., 2021).

## 2.2 | vis-NIR and mid-IR spectroscopy

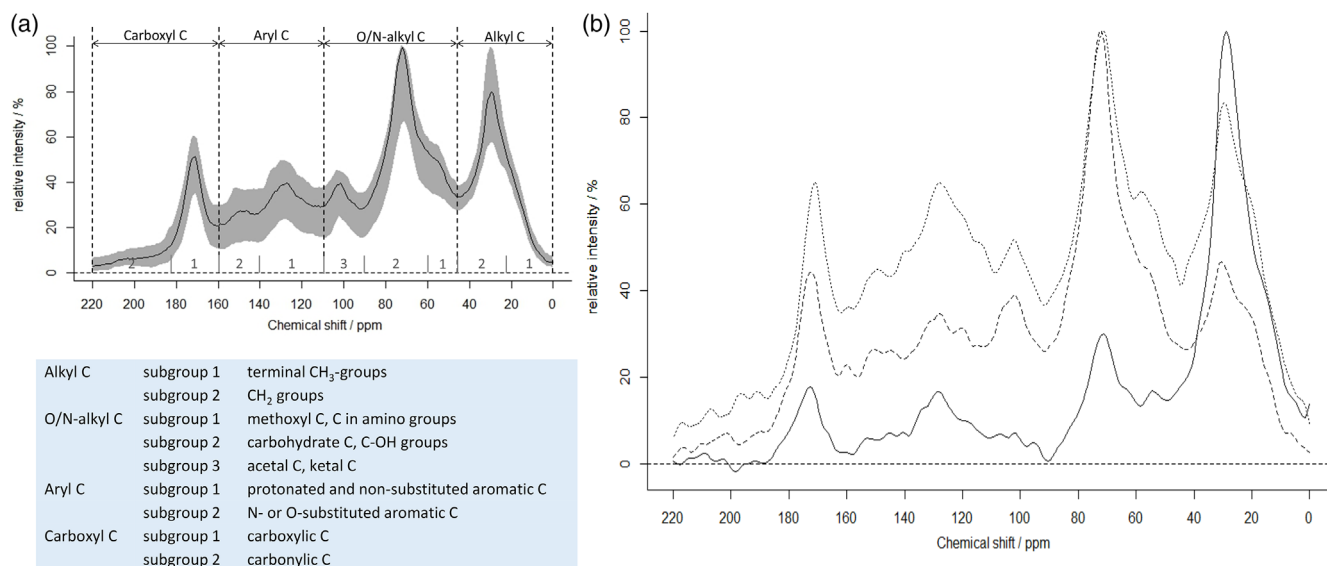
vis-NIR Spectra (350–2500 nm; 28,571 to 4000  $\text{cm}^{-1}$ ) were determined using an ASD FieldSpec Pro FR scanning instrument (Malvern Panalytical Ltd, Malvern, UK) on the 2 mm sieved and air-dried soil samples. The instrument was equipped with a bare optic fibre connected to a probe with a 20 W Al-coated halogen tungsten light source placed 7 cm over the sample, resulting in a field of view of approximately 7.5  $\text{cm}^2$ . Reflectance spectra were recorded in relation to an external white reference (Spectralon<sup>®</sup>) and each composite sample spectrum was comprised of 100 averaged spectra collected from a rotating sample. The spectra were sampled at 1.4–2 nm intervals with a spectral

resolution of 3–10 nm. A wavelength interval of 1 nm was interpolated to the instrument output file, resulting in spectra consisting of one data point for every nanometre. The vis-NIR spectra were transformed to apparent absorbance through  $\log(\text{reflectance}^{-1})$  and the 350–400 nm wavelength range was removed from further analysis due to noise.

Mid-IR spectra were recorded on four ground (<0.5 mm) replicates of each sample using an FT-IR Vertex70 spectrometer (Bruker, Germany) with a spectral range of 1333–16,667  $\text{nm}$  (7500–600  $\text{cm}^{-1}$ ) and a spectral resolution of 4  $\text{cm}^{-1}$  and 64 measurements per minute. The spectrometer was equipped with a nitrogen gas purging system to reduce the amount of atmospheric interference in the system which reduces masking of weak spectral features by water vapour or carbon dioxide absorption. A gold standard was used as reference. The mid-IR spectra were transformed to apparent absorbance through  $\log(\text{reflectance}^{-1})$  and only the 4000–600  $\text{cm}^{-1}$  (2500–16,700  $\text{nm}$ ) range was used in the further analysis. The four replicates were averaged to one spectrum per sample.

## 2.3 | <sup>13</sup>C NMR spectroscopy

The sieved samples were ground <0.630 mm using mortar and pestle prior to solid-state <sup>13</sup>C NMR experiments



**FIGURE 2** Summary of the NMR spectra from the 99 soil samples with (a) showing the median and 16th and 84th percentile NMR spectrum with the different functional C groups and subgroups indicated, and (b) showing NMR spectra representing three common types of NMR spectra in the data set, that is, those with a high, low and intermediate ratio of alkyl C to O/N-alkyl C

(Bruker DSX 200 NMR spectrometer, Karlsruhe, Germany) which were conducted at the NMR facility of the Institute of Soil Science at the Technical University of Munich. No paramagnetic material was present in the soils and consequently, no HF-treatment was required. The cross-polarisation magic angle spinning (CPMAS) technique was applied with a  $^{13}\text{C}$ -resonance frequency of 50.32 MHz and a spinning speed of 5 kHz. A ramped  $^1\text{H}$ -pulse was used during a contact time of 1 ms in order to circumvent spin modulation during the Hartmann-Hahn contact. A pulse delay of 1 s was used for all experiments and pre-experiments confirmed that the pulse delays were long enough to avoid saturation. Depending on the C contents of the samples, between 8000 and 400,000 scans were accumulated and a line broadening of 50 Hz was applied. The  $^{13}\text{C}$  chemical shifts were calibrated relative to tetramethylsilane (0 ppm).

Relative contributions of the various functional C groups were determined by integration of the signal intensity in their respective chemical shift regions according to Knicker et al. (2005). The region from 220 to 160 ppm was assigned to carbonyl (aldehyde and ketone) and carboxyl/amide C. Olefinic and aromatic C were detected between 160 and 110 ppm. O-alkyl and N-alkyl-C signals were found from 110 to 60 ppm and from 60 to 45 ppm. Resonances of alkyl C were assigned to the region 45 to 10 ppm. (Figure 2).

As indicator of the degree of decomposition of the SOC, the alkyl C:O/N-alkyl C ratio (45 to –10 ppm)/(110 to 45 ppm) was calculated from the NMR spectra (Baldock et al., 2004).

Correlations between the SOC, relative contribution of the different C groups derived from the  $^{13}\text{C}$  NMR spectra and the Alkyl C:O/N-alkyl C ratio derived from relative contribution were calculated using Spearman correlations.

## 2.4 | 2-D correlations

The raw NMR spectra were cut to only include the chemical shifts between 0 and 220 ppm where most of the information is found. The resolution of the three types of spectra was reduced to every 7.5th nm for the apparent absorbance vis-NIR spectra, every 10th  $\text{cm}^{-1}$  for the apparent absorbance mid-IR spectra, and every 0.8th ppm for the raw NMR spectra resulting in about 300 observations for all three spectral types. This was done to reduce the number of variables in the correlation analysis and to have a similar number of variables in all three spectra. The NMR spectra were further smoothed by a spline function and baselined using a second-order polynomial. The raw NMR spectra were then recalculated as relative intensity (relative to the most intense peak). Due to the shape of the mid-IR spectra these were first split into 4 regions (600–2100, 2100–2700, 2700–3720, 3720–4000  $\text{cm}^{-1}$ ) and then a baseline was applied using first, second, third, or fourth order polynomials to the different sections. After baselining the four sections were again recombined into one spectrum. To baseline the vis-NIR spectra we applied a continuum removal (Clark & Roush, 1984). The baselining was done to

further highlight and define the spectral features in the different spectra. A number of different baselining and smoothing techniques were tested, and the methods providing the visually best baselined spectra without artefacts were selected. The preprocessing of the spectra was performed in the statistical software environment R (R Core Team, 2020) and the hyperSpec (Beleites & Sergo, 2020) packages. The vis-NIR and mid-IR spectra were correlated to the NMR spectra by heterospectral correlation using the 2Dshige software (2Dshige@Shigeaki Morita, Kwansei-Gakuin University, 2004–2005). The correlations were plotted in 2-D plots for interpretation.

## 2.5 | Modelling functional C groups

The original apparent absorbance vis-NIR and mid-IR spectra were transformed and smoothed using first-order Savitzky–Golay derivative with 11 smoothing points (Savitzky & Golay, 1964). First-order derivative is a well-established pre-processing method in diffuse reflectance spectroscopy studies (Stenberg et al., 2010). A range of smoothing points was tested on a subset of SOC variables and the number producing the best cross-validated results was used in the final modelling. The vis-NIR and the mid-IR spectra were calibrated to the different NMR-derived functional C groups using support vector machines (SVM) with a radial basis function (RBF) kernel (Karatzoglou et al., 2006). Kernel-based learning methods use an implicit mapping of the input data into a higher dimensional feature space defined by a kernel function. With this, it is possible to derive a linear hyperplane as a decision function for non-linear problems (Vapnik, 1995). Here, we used a Gaussian RBF implemented in the kernlab library of the software R. Upper and lower bounds for the optimisation of the hyperparameters, penalty (C) and sigma of the RBF were set to 0 and 10, and 0 and 1 for C and sigma, respectively. The upper and lower bounds of the C and sigma parameters were used in the caret train function in the R library caret (Kuhn, 2008) and were optimised using the Differential Evolution optimisation (Price et al., 2006), implemented in the R library DEoptim (Mullen et al., 2011).

The models were validated using 10-fold (random) cross-validation repeated five times using the implementation in the caret library. The aggregation of the repeated cross-validations generates results that are more stable and robust. Thus we report the validation statistics and variable importance on the average of the five repeats. The validations were evaluated using the adjusted coefficient of determination ( $R^2$ ) of the linear relation between the predicted and measured values, the

concordance correlation coefficient (CCC), mean error (ME), the root-mean-square error (RMSE), which is a measure of the inaccuracy of the estimates and encompasses both bias and imprecision (Viscarra Rossel & McBratney, 1998). The CCC combines measures of both precision and accuracy (bias) and is calculated as

$$\frac{2r\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\mu_o - \mu_p)^2}$$

where  $r$  is the correlation coefficient between observed  $o$  and predicted  $p$ ,  $\mu_o$  and  $\mu_p$  are the means, and  $\sigma_o^2$  and  $\sigma_p^2$  are the corresponding variances.

To interpret the models, we calculated their variable importance using the varImp function in the caret library (Kuhn, 2008) of R.

## 3 | RESULTS

### 3.1 | Chemical composition of SOC

The 99 soil samples used in this study were selected from a total of a little over 12,500 Swedish arable topsoils to cover a large variation in SOC content and soil texture (Figure 1).

SOC varied from 1.3% to 10% and clay and sand content ranged from 5% to 40% and 80%, respectively. Because soil texture, and particularly clay content, has a significant influence on the vis-NIR and mid-IR spectra, we chose to use a data set without correlations between clay content and SOC and thus ensure independence in our analysis of SOC and its chemical composition.

Although all samples were collected from arable fields primarily under cereal crops, the SOC composition of the soil samples was variable, as shown by the different functional carbon groups defined with  $^{13}\text{C}$  NMR (Figure 2 and Table S2).

On average, the O/N-alkyl C group showed the largest contribution to the NMR spectrum. However, the alkyl C group with an average contribution of around 25% of the SOC showed the largest variation, contributing to up to 58% of the carbon in one sample. The NMR spectra of the soils could be classified into roughly three types depending on the contribution of the alkyl C and O/N-alkyl C groups: soil with a fairly even contribution from the two groups and soil with a primary contribution from either the alkyl C group or the O/N-alkyl C group (Figure 2b). The carboxyl-C group constituted the smallest portion of the total SOC. The degree of decomposition indicated by the ratio between the alkyl C and the O/N-alkyl C (Baldock et al., 2004) varied between 0.34 and 2.5 (Table S2).

The proportion of the alkyl C group increased with increasing SOC content ( $\rho = 0.76$ ,  $p < 0.05$ ) while the remaining carbon functional groups decreased (Table S3). The exception was the large O/N-alkyl C group that was not correlated with the total SOC content ( $\rho = -0.15$ ,  $p = 0.148$ ) and where the largest subgroup, the carbohydrates (O/N-alkyl C subgroup 2), was positively although fairly weakly, correlated to SOC ( $\rho = 0.45$ ,  $p < 0.05$ ). However, the acetal-ketal C portion of the O/N-alkyl C group (O/N-alkyl C subgroup 3) was negatively correlated to SOC, as were the other non-alkyl C groups. The strongest correlations occurred between the alkyl C group and the other functional groups. As the alkyl C group increased, the other functional groups made up smaller portions of the SOC. However, as for the correlations with SOC content, the largest O/N-alkyl C group, (O/N-alkyl C subgroup 2), was only weakly correlated with the alkyl C group ( $\rho = 0.22$ ,  $p < 0.05$ ). The ratio of alkyl C to O/N-alkyl C was strongly correlated to the alkyl C group and less so with the O/N-alkyl group ( $\rho = 0.98$ ,  $p < 0.05$ , and  $\rho = -0.66$ ,  $p < 0.05$ , for alkyl C and O/N-alkyl C groups, respectively).

### 3.2 | 2D correlations of $^{13}\text{C}$ NMR to diffuse reflectance spectra and modelling

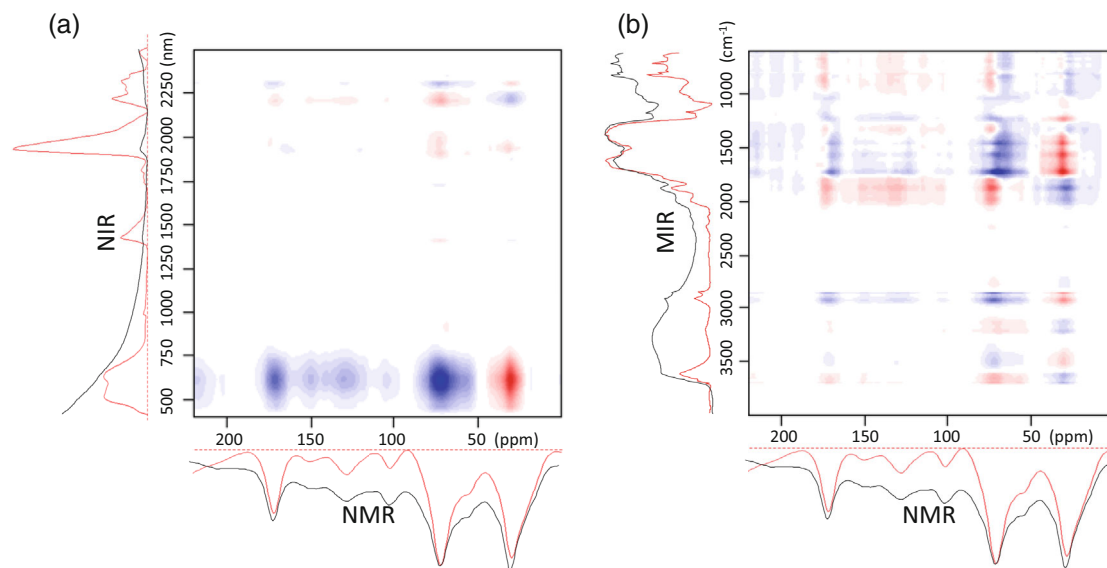
The relationship between the infrared spectra in the vis-NIR and mid-IR regions and the NMR spectra are shown in the 2D correlation plots, Figure 3.

The figure reveals general correlations between the functional C groups in the NMR spectra and different frequencies in the diffuse reflectance spectra. The vis-NIR spectra show the strongest positive correlations with the alkyl C group in the visible part of the spectrum with some weaker positive correlations around 2000 nm and 2300 nm (Figure 3a). The correlations to the remaining functional C groups show an opposite pattern to the correlations with the alkyl C group. The exception is a weak positive correlation with the O/N-alkyl C subgroup 2 at around 2200 nm.

The correlations between the NMR spectra and the mid-IR spectra were more detailed and less concentrated in one region of the spectrum. Although, similar to the correlations between NMR and the vis-NIR spectra, the general pattern show opposite correlations between the mid-IR spectra and the alkyl C group compared with the correlations between the mid-IR spectra and the other functional C groups (Figure 3b). Strong positive correlations with the alkyl C group occur in the 2800–3000  $\text{cm}^{-1}$  and the 1300–1700  $\text{cm}^{-1}$  regions (Figure 3b).

The diffuse reflectance spectra in the vis-NIR and the mid-IR regions were then used, individually, to model the different NMR-derived C functional C groups using SVM (Tables 1 and 2).

The best models with both the vis-NIR and mid-IR spectral regions was the alkyl C group, as a whole and especially the largest subgroup with  $\text{CH}_2\text{-C}$  (alkyl C subgroup 2) with adjusted  $R^2$  of 0.84 and 0.92 for vis-NIR and mid-IR models, respectively (Tables 1 and 2; Figure 4).



**FIGURE 3** 2D correlation plots between (a) vis-NIR and NMR, and (b) mid-IR and NMR. Blue colour indicates negative correlations and red indicates positive correlations. Black spectra show the average original absorbance IR spectra and relative intensity NMR spectra, and red the average base-lined spectra used in the correlations

	R <sup>2</sup>	CCC	RMSE	ME	Sigma	C
SOC	0.62	0.75	1.3	-0.207	0.00052	8.735
Alkyl C	0.83	0.91	3.0	0.090	0.00112	9.085
subgroup 1	0.56	0.74	1.6	0.087	0.00074	8.746
subgroup 2	0.84	0.91	2.0	-0.006	0.00110	8.199
O/N-alkyl C	0.34	0.56	2.6	-0.037	0.00187	8.479
subgroup 1	0.10	0.28	0.6	-0.014	0.00254	2.321
subgroup 2	0.26	0.44	1.7	0.027	0.00581	6.312
subgroup 3	0.70	0.82	0.6	0.026	0.00096	7.075
Aryl C	0.77	0.87	1.8	-0.144	0.00109	9.105
subgroup 1	0.73	0.84	1.2	-0.116	0.00111	7.355
subgroup 2	0.71	0.83	0.9	-0.002	0.00184	5.099
Carboxyl C	0.64	0.78	1.4	0.034	0.00123	8.531
subgroup 1	0.64	0.79	0.8	0.012	0.00116	7.466
subgroup 2	0.46	0.65	0.9	0.008	0.00552	3.535
Alkyl C: O/N-alkyl C	0.81	0.89	0.1	-0.007	0.00186	7.573

Note: The predictions are average results of five times repeated 10-fold cross-validation.

	R <sup>2</sup>	CCC	RMSE	ME	Sigma	C
SOC	0.86	0.92	0.8	-0.02	0.00053	5.841
Alkyl C	0.85	0.92	2.7	0.35	0.00048	9.024
subgroup 1	0.56	0.73	1.5	0.088	0.00027	9.137
subgroup 2	0.92	0.96	1.5	-0.017	0.00031	8.961
O/N-alkyl C	0.38	0.60	2.2	-0.003	0.00035	8.577
subgroup 1	0.32	0.54	0.5	0.017	0.00056	6.117
subgroup 2	0.20	0.42	1.7	-0.001	0.00075	7.366
subgroup 3	0.73	0.85	0.6	-0.054	0.00034	9.372
Aryl C	0.72	0.85	2.1	-0.098	0.00082	8.768
subgroup 1	0.71	0.83	1.3	-0.133	0.00039	5.835
subgroup 2	0.81	0.90	0.7	0.049	0.00062	9.955
Carboxyl C	0.54	0.73	1.5	0.006	0.00048	8.188
subgroup 1	0.63	0.77	0.7	-0.026	0.00025	9.126
subgroup 2	0.35	0.56	0.9	-0.10	0.00056	5.320
Alkyl C: O/N-alkyl C	0.84	0.91	0.1	0.00	0.00051	9.138

Note: The predictions are average results of five times repeated 10-fold cross-validation.

The importance of different wavelengths in the machine learning models (Figure 5) also shows clear contributions from the spectral regions corresponding to the asymmetric and symmetric CH-vibrations at 2930 cm<sup>-1</sup> and 2850 cm<sup>-1</sup>, respectively, in the mid-IR region, and their combination bands in the NIR around 2300 nm (Vicarria Rossel & Behrens, 2010). This corresponds to results shown in the 2D-correlation, although in the 2D-correlation plot of the NMR to mid-IR spectra, the alkyl-C group also showed positive correlations with the

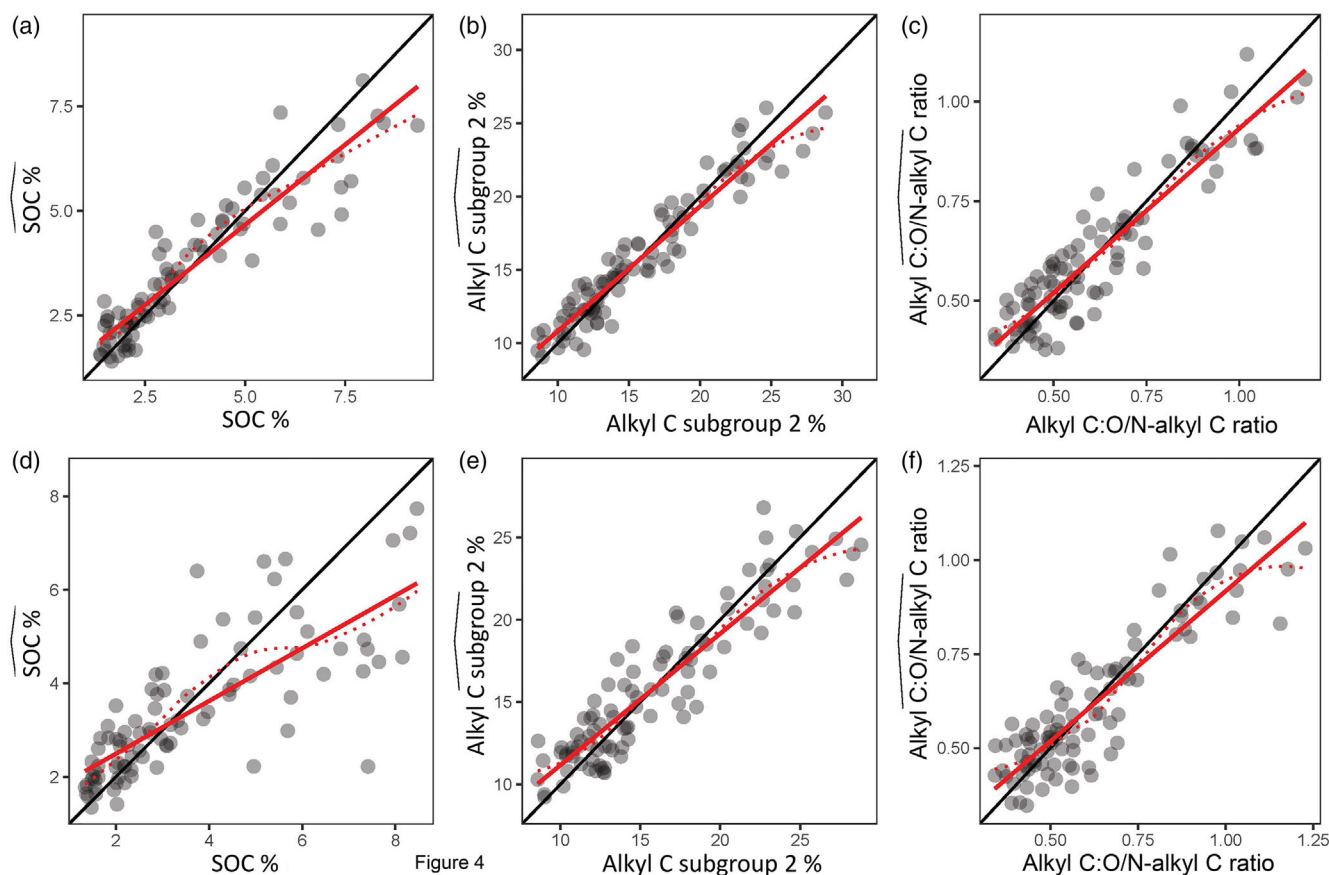
broader absorptions at 1700–1300 cm<sup>-1</sup> (Figure 3a) and the highest correlation between the vis-NIR and the alkyl-C group was actually in the visible region (Figure 3b).

The aryl C group was the second-best predicted functional C group with both vis-NIR and mid-IR models. Absorption near 1500 and 1700–1800 cm<sup>-1</sup> was important for prediction. Absorptions at 1500 cm<sup>-1</sup> can be attributed to aromatic C=C stretching vibrations and those near 1700 cm<sup>-1</sup> to C=O stretching vibrations (Tinti

TABLE 1 Cross-validated prediction results for the vis-NIR calibration models, and the final hyperparameters used (sigma, C), for SOC (%), relative contribution of the different <sup>13</sup>C NMR-derived C groups (%), and the alkyl C:O/N-alkyl C ratio derived from relative contribution

TABLE 2 Cross-validated prediction results for the mid-IR calibration models, and the final hyperparameters used, for SOC (%), relative contribution of the different <sup>13</sup>C NMR-derived C groups (%), and the alkyl C:O/N-alkyl C ratio derived from relative contribution





**FIGURE 4** Cross validated predictions versus measured SOC content (a), (d), relative contribution of alkyl C  $\text{CH}_2\text{-C}$  groups (alkyl C subgroup 2) (b), (e), and alkyl C:O/N-alkyl C ratio derived from relative contribution (c), (f) using machine learning models based on mid-IR (a)–(c) and vis-NIR (d)–(f) spectra. Red dotted lines are polynomial fits that show deviations from the linear fits. The predictions are the average of the five repeated cross-validations

et al., 2015). The 2D-correlation between the NMR and mid-IR also showed weak positive correlations between the aryl C group and the broad absorptions between 1700 and 1300  $\text{cm}^{-1}$ .

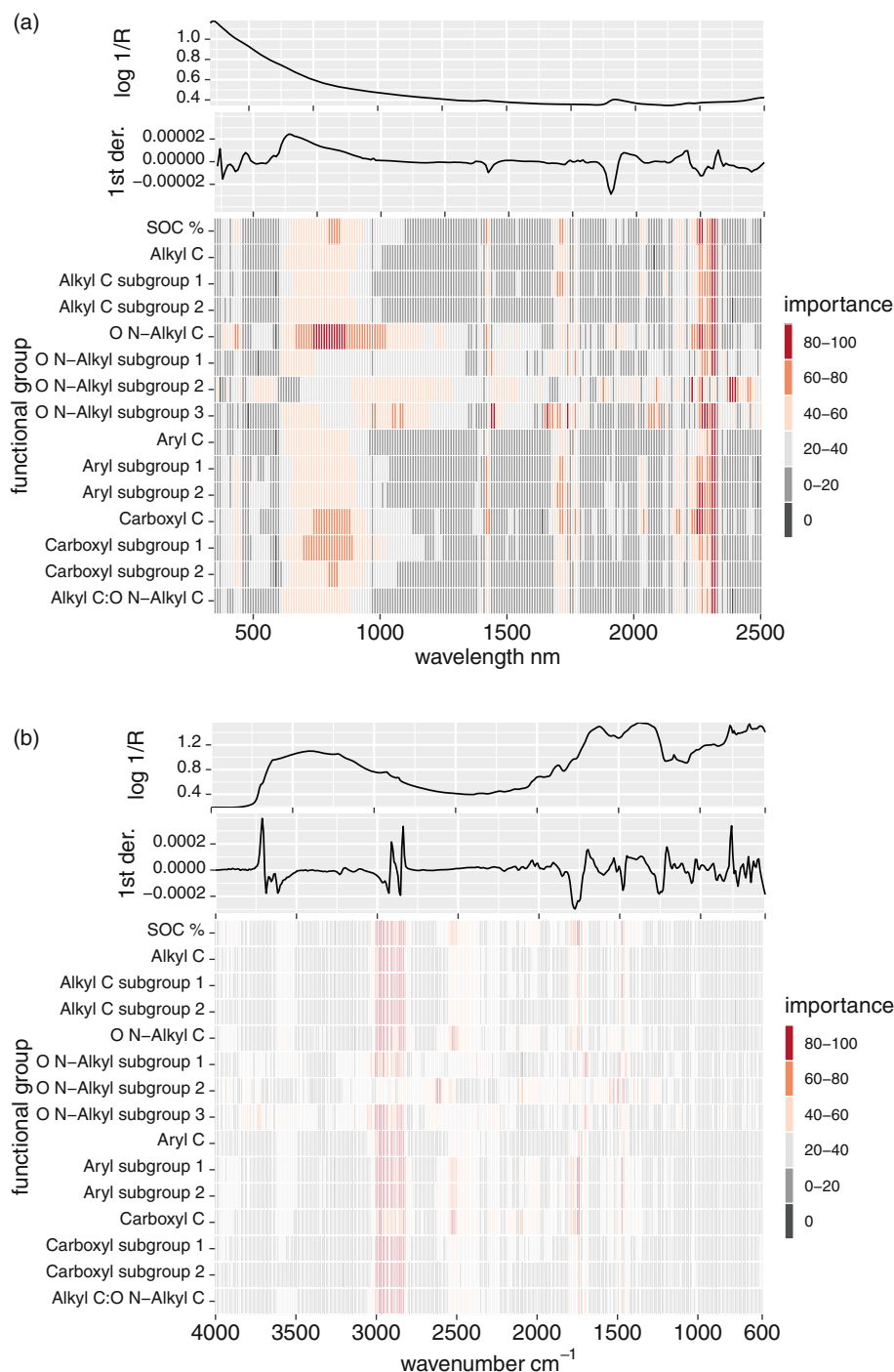
In the carboxyl C group, predictions of the well-defined subgroup 1 produced  $R^2 = 0.63\text{--}0.64$  using both vis-NIR and mid-IR spectra. However, predictions of the carboxyl C subgroup 2 were poor ( $R^2 = 0.35\text{--}0.45$ ), using mid-IR and vis-NIR spectra respectively. There are regions in the mid-IR spectra (e.g., 1642–1569  $\text{cm}^{-1}$ ) that are attributed to carboxylates, amongst other organic components (Tinti et al., 2015). However, this was not shown in our models. Rather, the similar pattern to the alkyl-C suggests that the carboxyl C subgroups were modelled based on negative correlations with the alkyl-C group (Table S2), whereas the large carboxyl C group show more similarities with the Aryl C subgroups.

The most difficult C group to predict in these soils was the large O/N-alkyl C group including carbohydrate C and C in amino groups. However, predictions of the small O/N-alkyl C subgroup 3, representing acetal and

ketal C, produced an  $R^2$  of around 0.7 using both vis-NIR and mid-IR spectra. One of the explanations for the difficulties in predicting this large C group might be the small variation in this group in our dataset, compared to, for example, the alkyl C group.

The modelling of the alkyl C:O/N-alkyl C ratio with both vis-NIR and mid-IR spectra produced  $R^2$  values of 0.81–0.84, which were similar to the  $R^2$  of the alkyl C group (Tables 1 and 2; Figure 4c and f). This was unsurprising because of the good predictability of the alkyl C and the very large variation in this C group compared with the O/N-alkyl C group.

The wavelength regions around 2000 nm showing weak positive correlations to the alkyl C groups and around 2200 nm showing weak positive correlations with the O/N alkyl C subgroup 2 in the 2D correlation plot have been reported to be important for OC modelling using vis-NIR (Stenberg et al., 2010), however, did not show as important in any of the models in this study. Absorbance at 2033 nm can be attributed to C=O vibrations (Viscarra Rossel & Behrens, 2010). Absorbance



**FIGURE 5** The importance of different wavelengths in the machine learning models for total C content, relative contribution of the different NMR-derived functional carbon groups and subgroups, and the ratio between alkyl C and O/N-alkyl C derived from relative contribution using the (a) vis-NIR and (b) mid-IR wavelength range

around 2200 nm is largely affected by minerals, for example, illite, which is a dominating mineral in the soils in this study (Stenberg et al., 2010).

Overall, models of the NMR-derived C functional groups using mid-IR were slightly better than those using vis-NIR spectra. However, the differences were not always large. The largest difference in the performance of the mid-IR and vis-NIR models was for total SOC content (Tables 1 and 2; Figure 4). The mid-IR model produced estimates of SOC that were as precise as the estimates for alkyl C

( $R^2 = 0.86$  for SOC compared with 0.85 for alkyl C). However, the estimates of SOC from the vis-NIR model were less precise ( $R^2 = 0.62$ ) than the estimates of alkyl C and aryl C (Table 1 and Figures 4a and d).

## 4 | DISCUSSION

The results provide further evidence that diffuse reflectance spectroscopy in the visible and infrared can be

used to estimate the chemical composition of SOC derived from  $^{13}\text{C}$  NMR in mineral bulk soil samples. The results presented in this study are based on young soils formed from quaternary deposits without paramagnetic material and with similar land management (arable fields), although presenting a large variation in climatic conditions, SOC content and soil texture. More studies, including other soil types and land management strategies, are needed to further evaluate under what conditions the methods could be used. Our results also demonstrate that spectroscopic estimates of SOC are soundly based on its chemical composition.

Our analyses used two approaches for relating the functional C groups to the vis-NIR and mid-IR spectra. First, using 2D heterospectral correlations between  $^{13}\text{C}$  NMR and infrared spectra, and second, using spectroscopic models of the specific functional C groups, which were derived from the  $^{13}\text{C}$  NMR. There was good correspondence in the results from the two methods, which strengthens our confidence in the findings.

The 2D-correlations showed the general associations between the  $^{13}\text{C}$  NMR and vis-NIR, and mid-IR spectra. The interpretation of the variable importance of the spectroscopic (vis-NIR and mid-IR) models of the functional C groups were similar but they revealed more detail. For the different C-groups, many of the important wavelength regions in the models were similar (Figure 5), but there were some notable differences, for example, comparing the mid-IR models of aryl C and alkyl C (Figure 5b). This also suggests that the chemical composition of SOC can be characterised separately, and is not based on SOC content. However, some of the C-groups seem to be modelled largely based on indirect correlations with other C groups which have a negative effect on model robustness.

Generally, models using mid-IR spectra produced better estimates compared to vis-NIR models. This is because the fundamental vibrations occur in the mid-IR region whereas the NIR spectra result from overtones and combinations of these vibrations (Soriano-Disla et al., 2014). However, the differences were often small. One reason for this could be the contribution of the visible range to the vis-NIR models. The visible region shows the strongest correlations with the NMR-spectra in the 2D-correlation plot (Figure 3a) and the visible and short-wave NIR regions (<1000 nm) are indicated as important regions in the models (Figure 5a). The response in the visible region due to organic matter is broad but clear, and several studies have reported the improved modelling of SOC when the visible and the NIR regions are combined (Stenberg et al., 2010). The advantage of using mid-IR compared to vis-NIR seems to

be in the estimation of total SOC in datasets with a large variation in the composition of SOC. The estimates of SOC using vis-NIR spectra appear to be better at smaller SOC concentrations but deteriorate at SOC contents above 4% (Figure 4d). Ben-Dor and Banin (1995) found similar problems with using NIR spectroscopy to estimate soil organic matter in a data set with variable degrees of decomposition of the organic matter depending on organic matter content. We found a clear correlation between SOC and C composition in the soils used in our study, with an increase in the proportion of alkyl C with increasing SOC content, but also an increased variation in the proportion of the alkyl C with an increase in SOC (data not shown).

The soil samples used in this study are mineral agricultural soils. The diversity of the C inputs is narrow and, as might be expected, so is the variability of the SOC. Nonetheless, the samples originate from a large geographic extent, covering different climatic regions and with diverse soil textures (Figure 1), which introduces variability in decomposition conditions of the soils used. Apart from the O/N-alkyl C group that constituted a smaller portion of the total SOC and was less variable in our study, the proportions and ranges of the functional C groups were similar to those of studies with more diverse samples, including forest litter, specific soil fractions and soils from different land uses (Leifeld, 2006; Terhoeven-Urselmans et al., 2006).

The promising but somewhat inconsistent results in the few studies published on this subject (e.g., Leifeld, 2006; Terhoeven-Urselmans et al., 2006; 431 Ludwig et al., 2008; Forouzangohar et al., 2015; Kang et al., 2017) may be attributed to the large variability within the samples, both between and within studies, and the often small number of samples used in those studies. Other studies Terhoeven-Urselmans et al., 2006; Ludwig et al., 2008) reported better estimates of O/N alkyl C than our study. This might be due to the relatively small variation in O/N alkyl C in our study (23%–50%) compared to those studies which included samples with more less-decomposed material leading to higher and more variable O/N alkyl C content (33%–82%). Differences in C inputs, with more diverse materials for example, including coniferous materials in many of the published studies, might also partly explain the differences in the accuracy of the aryl C and carboxyl C group estimates. The relatively more homogeneous C inputs and SOC of the sample set in our study might have contributed to the better estimates of the alkyl C:O/N-alkyl C ratio.

The use of HF-treated soils in some of the other studies (Forouzangohar et al., 2013, 2015) prevents direct comparisons to our results. However, our results are encouraging because we obtained good estimates of the alkyl C and alkyl C:O/N-alkyl C ratio in whole mineral

soils without any fractionation or chemical pretreatments ( $R^2$  for Alkyl C = 0.83 and 0.85, and  $R^2$  for alkyl C:O/N-alkyl C ratio = 0.81 and 0.84, for vis-NIR and mid-IR respectively). No paramagnetic material was present in the soils in this study and the results are valid for soils under similar conditions.

## 5 | CONCLUSIONS

The study shows that diffuse reflectance spectroscopy in the visible and infrared can be used to estimate the chemical composition of SOC in whole mineral soil samples without C fractionation or HF-treatment. The results further demonstrate that spectroscopic estimates of SOC are soundly based on its chemical composition.

Although diffuse reflectance spectroscopy may not estimate SOC composition as accurately as  $^{13}\text{C}$  NMR, and there is still a need for traditional methods for calibrations, the opportunity to analyse more samples due to the more cost-efficient analysis could improve the detection and monitoring of changes that might otherwise be lost due to spatial variation. Diffuse reflectance spectroscopy also enables in-field measurements, which make it possible to consider in-situ measurements of SOC composition from soil that is under field condition and undergoing decomposition.

### AUTHOR CONTRIBUTIONS

**Johanna Wetterlind:** Data curation (lead); formal analysis (equal); funding acquisition (lead); investigation (lead); methodology (equal); writing – original draft (lead); writing – review and editing (equal). **Raphael A. Viscarra Rossel:** Funding acquisition (supporting); investigation (supporting); methodology (equal); writing – original draft (supporting); writing – review and editing (equal). **Markus Steffens:** Investigation (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting).

### ACKNOWLEDGEMENTS

This project was funded by the Swedish Research Council Formas (Dnr. 229-2010-951 and Dnr. 2017-00887) RAVR's contribution was supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP210100420). The authors are grateful to Ingrid Kögel-Knabner for enabling the extensive soilid state  $^{13}\text{C}$  NMR experiments at her institution.

### DATA AVAILABILITY STATEMENT

The data used are available from the corresponding author on reasonable request.

### ORCID

Johanna Wetterlind  <https://orcid.org/0000-0003-2821-2999>

Raphael A. Viscarra Rossel  <https://orcid.org/0000-0003-1540-4748>

### REFERENCES

- Audette, Y., Congreves, K., Schneider, K., Zaro, G. C., Nunes, A. L. P., Zhang, H., & Voroney, R. P. (2021). The effect of agroecosystem management on the distribution of c functional groups in soil organic matter: A review. *Biology and Fertility of Soils*, *57*, 881–894.
- Baldock, J., Masiello, C., Gelin, Y., & Hedges, J. (2004). Evaluation of spectral pretreatments, spectral range, and regression methods for quantitative spectroscopic analysis of soil organic carbon composition. *Spectroscopy Letters*, *92*(1–4), 39–64.
- Baldock, J., Oades, J., Nelson, P., Skene, T., Golchin, A., & Clarke, P. (1997). Assessing the extent of decomposition of natural organic materials using solid-state  $^{13}\text{C}$  nmr spectroscopy. *Australian Journal of Soil Research*, *35*, 1061–1083.
- Baldock, J. A., Oades, J. M., Vassallo, A. M., & Wilson, M. A. (1989). Incorporation of uniformly labelled  $^{13}\text{C}$ -glucose carbon into the organic fraction of a soil. Carbon balance and CP/MAS  $^{13}\text{C}$  NMR measurements. *Australian Journal of Soil Research*, *27*, 725–746.
- Baldock, J. A., Oades, J. M., Waters, A. G., Peng, X., Vassallo, A. M., & Wilson, M. A. (1992). Aspects of the chemical structure of soil organic materials as revealed by solid-state  $^{13}\text{C}$  nmr spectroscopy. *Biogeochemistry*, *16*(1), 1–42.
- Beleites, C. & Sergio, V. (2020). hyperSpec: a package to handle hyperspectral data sets in R. R package version 0.99-20201127. <https://github.com/cbeleites/hyperSpec>
- Ben-Dor, E., & Banin, A. (1995). Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, *59*(2), 364–372.
- Bonanomi, G., Incerti, G., Giannino, F., Mingo, A., Lanzotti, V., & Mazzoleni, S. (2013). Litter quality assessed by solid state C-13 NMR spectroscopy predicts decay rate better than CN and LigninN ratios. *Soil Biology & Biochemistry*, *56*, 40–48.
- Clark, R., & Roush, T. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, *89*, 6329–6340.
- Ekström, G. (1927). Klassifikation av Svenska Åkerjordar (Classification of Swedish arable soils). Tech. Rep. 345 (Årsbok 20), Sveriges Geologiska Undersökning.
- Forouzanoghar, M., Baldock, J., Smernik, R., Hawke, B., & Bennett, L. (2015). Mid-infrared spectra predict nuclear magnetic resonance spectra of soil carbon. *Geoderma*, *247*, 65–72.
- Forouzanoghar, M., Cozzolino, D., Smernik, R., Baldock, J., Forrester, S., Chittleborough, D., & Kookana, R. (2013). Using the power of C-13 NMR to interpret infrared spectra of soil organic matter: A two-dimensional correlation spectroscopy approach. *Vibrational Spectroscopy*, *66*, 76–82.
- Gee, G. W., & Bauder, J. W. (1986). Particle-size analysis. In A. Klute (Ed.), *Physical and mineralogical methods* (2nd ed., pp. 383–411). Soil Science Society of America.
- Kang, H., Gao, H., & Yu, W. (2017). Evaluation of spectral pretreatments, spectral range, and regression methods for quantitative

- spectroscopic analysis of soil organic carbon composition. *Spectroscopy Letters*, 50(3), 143–149.
- Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software*, Articles, 15(9), 1–28.
- Karlsson, C., Sohlenius, G., & Peterson Becher, G. (2021). *Handleddning för jordartsgeologiska kartor och databaser över Sverige*, SGU-rapport 2021:17. Sveriges Geologiska Undersökning (SGU).
- Kinchesh, P., Powlson, D. S., & Randall, E. W. (1995). C NMR studies of organic matter in whole soils: 13 I. Quantitation possibilities. *European Journal of Soil Science*, 46, 125–138.
- Knicker, H., Gonzalez-Vila, F., Polvillo, O., Gonzalez, J., & Almendros, G. (2005). Fire-induced transformation of c- and n-forms in different organic soil fractions from a dystric cambisol under a mediterranean pine forest (pinus pinaster). *Soil Biology & Biochemistry*, 37, 701–718.
- Knox, N. M., Grunwald, S., McDowell, M. L., Bruland, G. L., Myers, D. B., & Harris, W. G. (2015). Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma*, 239, 229–239.
- Kögel-Knabner, I. (1997). Analytical approaches for characterizing soil organic matter. *Geoderma*, 80(3–4), 243–270.
- Kögel-Knabner, I. (2000). Analytical approaches for characterizing soil organic matter. *Organic Geochemistry*, 31, 609–625.
- Kögel-Knabner, I., & Rumpel, C. (2018). Chapter one - advances in molecular approaches for understanding soil organic matter composition, origin, and turnover: A historical overview. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 149, pp. 1–48). Academic Press.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, Articles, 28(5), 1–26.
- Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, 528(7580), 60–68.
- Leifeld, J. (2006). Application of diffuse reflectance FT-IR spectroscopy and partial least-squares regression to predict NMR properties of soil organic matter. *European Journal of Soil Science*, 57(6), 846–857.
- Ludwig, B., Nitschke, R., Terhoeven-Urselmans, T., Michel, K., & Flessa, H. (2008). Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde*, 171(3), 384–391.
- Mathes, N., Xu, Z., Berners-Price, S., Perera, M., & Saffigna, P. (2002). Hydrofluoric acid pre-treatment for improving <sup>13</sup>C cpmas nmr spectral quality of forest soils in south-East Queensland, Australia. *Australian Journal of Soil Research*, 40, 655–674.
- Miller, C. E. (2001). Chemical principles of near-infrared technology. In P. Williams & K. Norris (Eds.), *Near-infrared Technology in the Agri-cultural and Food Industries*. Springer-Verlag New York, Inc..
- Mullen, K., Ardia, D., Gil, D. L., Windover, D., & Cline, J. (2011). Deoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6), 1–26.
- Oades, J. M., Vassallo, A. M., Waters, A. G., & Wilson, M. A. (1987). Characterization of organic matter in particle size and density fractions from a red-brown earth by solid state <sup>13</sup>C nmr. *Soil Research*, 25, 71–82.
- Paré, M. C., & Bedard-Haughn, A. (2013). Soil organic matter quality influences mineralization and ghg emissions in cryosols: A field-based study of sub-to high arctic. *Global Change Biology*, 19(4), 1126–1140.
- Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimization*. Springer Science & Business Media.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Sanderman, J., Farrell, M., Macreadie, P., Hayes, M., McGowan, J., & Baldock, J. (2017). Is demineralization with dilute hydrofluoric acid a viable method for isolating mineral stabilized soil organic matter? *Geoderma*, 304, 4–11.
- Savitzky, A., & Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627–1639.
- Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kogel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse, D. P., Weiner, S., & Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, 478, 49–56.
- Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., & McLaughlin, M. J. (2014). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49(2), 139–186.
- Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107, 163–215.
- Terhoeven-Urselmans, T., Michel, K., Helfrich, M., Flessa, H., & Ludwig, B. (2006). Near-infrared spectroscopy can predict the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde*, 169(2), 168–174.
- Tinti, A., Tugnoli, V., Bonora, S., & Francioso, O. (2015). Recent applications of vibrational mid-infrared (IR) spectroscopy for studying soil components: A review. *Journal of Central European Agriculture*, 16(1), 1–22.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc.
- Viscarra Rossel, R., & McBratney, A. (1998). Soil chemical analytical accuracy and costs: Implications from precision agriculture. *Australian Journal of Experimental Agriculture*, 38(7), 765–775.
- Viscarra Rossel, R. A., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1–2), 46–54.
- Viscarra Rossel, R. A., & Hicks, W. S. (2015). Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. *European Journal of Soil Science*, 66, 438–450.
- Viscarra Rossel, R. A., Lee, J., Behrens, T., Luo, Z., Baldock, J., & Richards, A. (2019). Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls. *Nature Geoscience*, 12(7), 547–552.
- Weng, Z., Lehmann, J., Zwieten, L. V., Joseph, S., Archanjo, B. S., Cowie, B., Thomsen, L., Tobin, M. J., Vongsivut, J., Klein, A., Doolette, C. L., Hou, H., Mueller, C. W., Lombi, E., &

Kopittke, P. M. (2021). Probing the nature of soil organic matter. *Critical Reviews in Environmental Science and Technology*. <https://doi.org/10.1080/10643389.2021.1980346>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wetterlind, J., Viscarra Rossel, R. A., & Steffens, M. (2022). Diffuse reflectance spectroscopy characterises the functional chemistry of soil organic carbon in agricultural soils. *European Journal of Soil Science*, 73(4), e13263. <https://doi.org/10.1111/ejss.13263>